



**HAL**  
open science

## Signal estimation and uncertainties extraction in TeraHertz Time Domain Spectroscopy

Elsa Denakpo, Théo Hannotte, Nouredin Osseiran, François Orioux, Romain  
Peretti

► **To cite this version:**

Elsa Denakpo, Théo Hannotte, Nouredin Osseiran, François Orioux, Romain Peretti. Signal estimation and uncertainties extraction in TeraHertz Time Domain Spectroscopy. 2024. hal-04725491v2

**HAL Id: hal-04725491**

**<https://cnrs.hal.science/hal-04725491v2>**

Preprint submitted on 10 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Signal estimation and uncertainties extraction in TeraHertz Time Domain Spectroscopy

Elsa Denakpo\*, Théo Hannotte\*, Nouredin Osseiran\*, François Orioux†, and Romain Peretti\*

**Abstract**—Terahertz Time Domain Spectroscopy (THz-TDS) systems have emerged as mature technologies with significant potential across various research fields and industries. However, the lack of standardized methods for signal and noise estimation and reduction hinders its full potential. This paper introduces a methodology to significantly reduce noise in THz-TDS time traces, providing a reliable and less biased estimation of the signal. The method results in an improved signal-to-noise ratio, enabling the utilization of the full dynamic range of such setups. Additionally, we investigate the estimation of the covariance matrix to quantify the uncertainties associated with the signal estimator. This matrix is essential for extracting accurate material parameters by normalizing the error function in the fitting process. Our approach addresses practical scenarios where the number of repeated measurements is limited compared to the sampling time axis length. We envision this work as the initial step toward standardizing THz-TDS data processing. We believe it will foster collaboration between the THz and signal processing communities, leading to the development of more sophisticated methods to tackle new challenges introduced by novel setups based on optoelectronic devices and dual-comb spectroscopy.

**Index Terms**—time domain spectroscopy, signal processing, terahertz (THz), noise extraction, covariance inverse estimation

## I. INTRODUCTION

Terahertz (THz) frequency range has long been a frontier for technology, too fast for electronics and at too low energy to ensure proper laser radiation. Therefore, THz spectroscopy was only rarely used and only for research purposes despite much interest in gas, liquid, and solid phase samples. The paradigm shifted a few decades ago thanks to the use of opto-electronics sources coupled to new generations of both continuous and pulse lasers. This enabled the performances of THz spectroscopy from a few fraction of THz to a few THz with a very good dynamic range using a bench-top setup. This highly contributed to the spread of THz spectroscopy, which is used to characterize the physical and chemical properties of samples from the solid, liquid, gaseous, and even plasma phases [1].

In gas, THz covers both the rotational transitions of the microwave regime for relatively small molecules (3 to 15 atoms) and the vibrational modes of the infrared for larger molecules. Therefore, it is known to be the most selective spectral range for organic volatile compounds [2] with potential application in atmospheric study, air pollution monitoring,

or breathalyzers. In liquids, water picosecond dynamics is in the same time an open fundamental subject and very sensitive to interaction with solutes and thus provides important information about proteins solvation [3, 4] for instance. Finally, in the solid state, it is used in many fundamental semiconductor [5–7], spintronic [8, 9], and 2D material research [10, 11], as a very good tool to monitor the crystal phase of organic molecules [12, 13]. THz is therefore very attractive for many analytical fields and is used to characterize tablets porosity during their fabrication in the pharmaceutical industry and in the automotive one - with the major example of measuring the thickness of paint layers during the painting of automotive parts or non-destructive analysis of thin film elements.

The setups and techniques have developed now and commercially available time-domain spectroscopy setups offer more than 100 dB of dynamics range and bandwidth of about 8 THz and are distributed and used by scientists and engineers all over the world. THz spectroscopy is a rapidly growing field with a wide range of applications. During the experiments the instrument records two time traces one without sample and one with sample. From there, several teams have proposed methodologies to extract useful information for the physicist, chemist or material scientist. Historically [14–17], the sequence of information extraction included, a FFT, phase unwrap, refractive index evaluation thanks to a basic and often valid photonic model, and finally the parameter estimation. They used either a better model for the recorded signal [18], or went further in classifying samples through data processing [19, 20]. The last approach was to present statistically valid estimation to extract the parameters when fitting the experimental data [21]. Still, despite these efforts [22], there is no standard in the data acquisition and data processing and no accepted protocol to extract the uncertainties from a measurement and therefore rare are the studies proposing error bars in their measurements.

In this work, we propose to go a step further in the data acquisition and processing in THz-TDS experiments thanks to a robust estimator of the signal on the one hand, and an estimation of the noise correlation matrix needed for materials parameters extraction on the other hand. The estimator is based on the average of the time traces corrected depending on the source of noise, by applying a generic optimization approach. The noise correlation matrix for its part is approximate with algorithms used for covariance and precision matrices estimation. Both are included in the open source free software Correct@TDS [23], which implements our estimator and enables the estimation of the noise correlation matrix. The manuscript first shortly describes the THz-TDS setup and the methods used in the community to process the data,

The authors marked with \* are with Univ. Lille, CNRS, Univ. Polytechnique Hauts-de-France, UMR 8520 IEMN - Institut d'Electronique de Microélectronique et de Nanotechnologie, F-59000 Lille, France, marked with † with Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France. Correspondance should be sent to romain.peretti@cnrs.fr

specifies the different sources of noise and ways to mitigate it, proposes an estimator for the signal and the corresponding noise correlation matrix. We believe that this approach and software will help the technique to spread more in analytical fields and hopefully to draw a bridge to the data processing and instrumentation community.

## II. TERAHERTZ TIME DOMAIN SPECTROSCOPY SETUP, DATA EXPLOITATION AND EXPERIMENTAL ERRORS

### A. TeraHertz Time Domain spectroscopy experiments

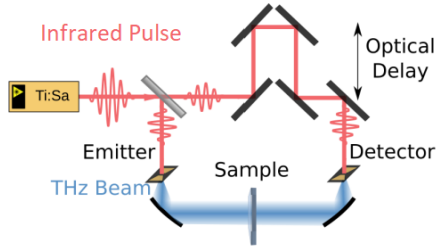


Fig. 1: THz-TDS experiment principle showing the main elements and source of errors through noises and perturbations: the delay line and the detector and the laser.

In a THz-TDS experiment (Fig. 1), a femtosecond laser emits an ultrafast pulse that is split in two by a beam splitter and directed to two photoconductive antennas. When the ultrashort pulse excites the transmitting antenna, it creates carriers which are accelerated by a high voltage between two electrodes and induce the terahertz pulse. The detectors works on a very similar principle and the delay line allows to adjust the length of the path before reaching the detection antenna. Thus by varying the time delay between the pulse on the detector and the one on the emitter, the E-field is measured as a function of time. At the end of an experiment, we obtain two time traces, one without sample that we call the reference and one with sample that we call sample. In both cases, the electric field measured versus time, which will be analyzed to retrieve information on the properties of the studied material.

### B. TeraHertz Time Domain spectroscopy First data process

From these time traces, several further steps have been proposed to extract information from the experiments. The first one is the computing of the Transmission; it consists on a simple performance of the ratio between the absolute value of the FFT of the time traces. The second one add a log to the previous step to get the loss per unit length when divided by the thickness of the sample. The third one is a bit more complex, it consists on the extraction of the complex refractive indices versus frequency, of the studied material. It is again usually done from the frequency domain by Fourier-transform of the recorded time traces. Then, a fitting process is used, to get the two real value describing the complex refractive index [24]. Finally, one can further extract information by fitting the dispersion of the refractive index by permittivity models to get physical insight about the material. This method in the frequency domain reach its limitations for

instance when the signal is too noisy due to a too strong absorption [25]. Indeed, the accuracy of the measured time traces have a direct influence on the performance of the fitting model. Moreover, the fit by the permittivity models usually do not take into account the noise. This omission would be valid if the noise would be a Gaussian white noise but in the experiments this assumption was actually never confirmed. To overcome this issue, several group decided to look back to the initial time traces in the performance of the data processing. Mohtashemi et al. [26] proposed a method that includes the noise correlation matrix which is a fair description of the uncertainty due to the noise in the reference and in the sample time traces. In our group [27] and [28, 29], we proposed a software called Fit@TDS which implements a method where the permittivity models are directly fitted from the data in time domain and include the noise correlation matrix. This two new method show the importance of looking deeper on the source of error from the systems in THz-TDS experiments.

### C. THz-TDS systems noises and perturbations

THz-TDS systems are now well matured but their capabilities can still be improved beside simple the output power of the emission antenna and the bandwidth of the systems. Indeed, the interpretability and the reproducibility of the output data is of utter importance specifically regarding application such as in biology [30]. This relies on a deep understanding of all the sources of error, including noises and perturbations, induce by the system itself. THz-TDS are known to have a very good dynamic range, but up to now there is no standard definition of the signal to noise ratio. Thus it's important to understand where the noise comes from the system. Many studies [31–36] had been done on the subject of noise sources in THz-TDS systems. Jahn et al. [37] described how the uncertainties in the delay line positioning influences the acquired signal and its spectrum. Rehn et al. [38] explained how the periodic error in delay line line position can make the spectrum unusable. Hence, the main sources of noise identified are the variation of the laser power and the delay line position which impact the time sampling step.

To better quantify and understand the noise added to the signal by the spectrometer, we performed a set of experiments. Our setup is a MenloSystems's terahertz spectrometer TeraSmart. This setup is relatively standard using a 100 MHz repetition rate laser and a 850 ps long delay line covering 8.5% of the laser period with a time step of 33 fs. The typical THz bandwidth is around 4 THz.

For the sake of comprehension of the rest of this work, we denote:

$$T_i = (E_i(t))_h \quad (1)$$

$$\tilde{T}_i = (\tilde{E}_i(f))_{f \in \{0, \dots, \lfloor \frac{p-1}{2} \rfloor\}} \quad (2)$$

$$\mathcal{T} = (T_i)_{i \in \{1, \dots, n\}} \quad (3)$$

$$\bar{T} = \left( \frac{1}{n} \sum_{i=1}^n E_i(t) \right)_{t \in \{t_0, \dots, t_{p-1}\}} \quad (4)$$

with  $T_i$  the  $i$ th measured time trace,  $\tilde{T}_i$  its Fourier transform,  $\mathcal{T}$  the set of  $n$  time traces measured repeatedly, and  $\bar{T}$  the average along the time axis of all time traces  $T_i, i \in \{1 \dots n\}$ .

For the first experiment the two optical fibers are unplugged and we recorded the signal on 845 ps without THz emission and activation of the detector. We call this noise "Dark noise" and it is generated by the detector when no THz pulse is sent to the detector. In fact, even when no carrier are promoted to the conduction band, some hole-electron pairs are created, generating a small current, that is amplified by the transimpedance amplifier, itself adding noise. Fig. 2 shows the average on 1000 single record in frequency domain. Its average power decreases when accumulating, it goes from  $1 \times 10^{-1} \mu\text{V}$  for an average on 30 time traces to  $1.1 \times 10^{-2} \mu\text{V}$  for an average of 3000 time traces. It is non deterministic and has the flat spectrum of a white noise.

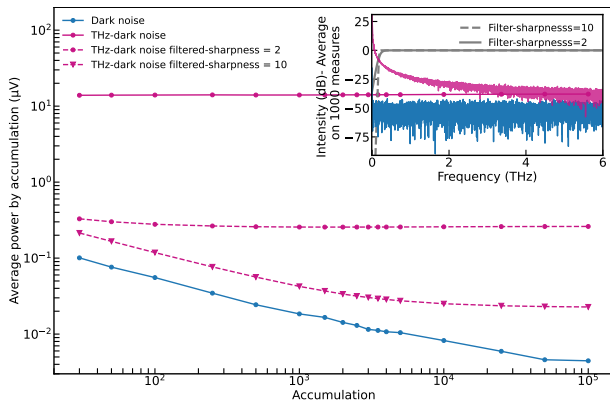


Fig. 2: Noise recorded in the different experimental conditions: Dark noise and THz-Dark noise with different filtering showing that this filtering is unavoidable to keep a reasonable noise even in the absence of signal.

A second experiment is to put an absorbent to block the cell occupied by the sample and record the signal on 845 ps. Both antennas are excited by the femtosecond laser and we call this "THz-Dark perturbation"; its amplitude as well as its shape include a low frequency artifact-perturbation that is reproducible [29]. As shown in Fig. 2, its amplitude is two orders of magnitude greater than the Dark noise recorded previously in our first experiment and its shape has a deterministic trend. Here, we thoroughly ensure that this noise was not due to parasitic THz reflection in our setup by translating and rotating the antenna without changing this perturbation. This implies that some parasitic laser emission reach the detector after the main pulse and after going through the delay line. At low frequency, its contribution is important and must be reduced. This was done thanks to a low cut filter at 200 GHz corresponding to the low bounds frequency of the bandwidth given by the manufacturer of our setup. The filter is a smooth step function and is defined by:

$$H(f) = 0.5 + 0.5 \tanh\left(\frac{(f - f_{cut})\beta}{f_{cut}}\right) \quad (5)$$

with  $f_{cut}$  the cut-off frequency and  $\beta$  the sharpness. Fig. 2 shows the average power of this noise. After filtering for an

accumulation of 3000 time traces the THz-Dark noise average power goes from  $14 \mu\text{V}$  before filtering to  $3 \times 10^{-2} \mu\text{V}$  after filtering for a sharpness of 10, similar to the Dark noise order of magnitude.

From this experiments, it appears that the signal should be filter at low frequencies in order to reduce both noises. Also the sharpness of the filter can have a great influence on the result. In fact, 200 GHz is the frequency limit of our system, and a large sharpness means more correlation added to the data. Hence a filter under 200 GHz with a sharpness of 2 is a good compromise to not impact the most important part of the signal at higher frequencies.

A third experiment is to record THz pulse traveling in dry nitrogen between the two antennas, plugged to the optical fibers. Here 100 000 time traces on 100 ps have been acquired consecutively. Fig. 3 shows the average on 1000 time traces and the standard error on 100 averages on 1000 time traces.

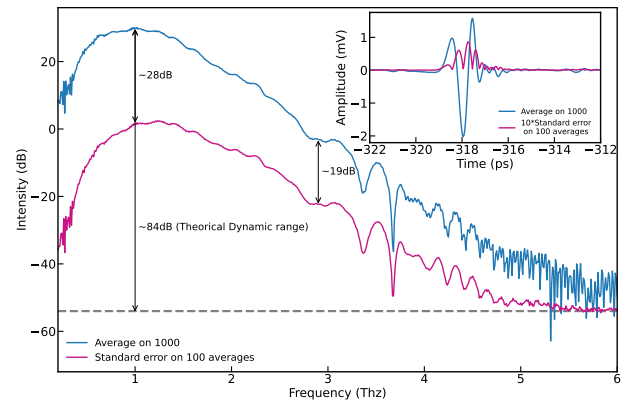


Fig. 3: Mean and standard error in frequency and time domain on time traces recorded in dry nitrogen showing that the noise cannot be said to be equal to neither the dynamic range nor to the dark noise.

When averaging on time traces samples, the goal is to increase the signal to noise ratio. The average is then considered as an estimator of the signal and the standard error it's standard deviation. Indeed, in statistics, when the samples are independent and identically distributed (i.i.d.) from a Gaussian distribution with mean  $\mu$  the average - also called the sample mean - is the maximum likelihood estimator of this mean  $\mu$ .

As we can see in Fig. 3, the standard error is correlated to the signal and the signal to noise ratio is about 28 dB while the dynamic range as define in the community (maximum of the signal compare to the high frequency noise) is  $\sim 84$  dB. The first conclusion is that for the estimator "average", the noise estimated by the standard error is very high compare to the dynamic range. Two causes can lead to this effect: There is a very important noise correlated to the signal; the estimator "average" on the raw data is not a good estimator for this experiments. In the next sections, we will propose to correct most of the noise coming from the system to build a better estimator of the signal and increase the signal to noise ratio.

### III. SIGNAL ESTIMATOR

#### A. Delay line initial position drift

While accumulating multiple measurements is a common approach to improve signal estimation, the simple average often results in a biased estimator for time-domain sampled signals as already noted in [39]. Specific efforts have been pursued to retrieve unbiased signal estimators in time domain experiments. In our case, upon closer examination of the data, we discovered that the time traces exhibited slight shifts of a few femtoseconds between each accumulation. It translates in a shape of the standard deviation of the time signal similar to the one shown in [40] and the inset of 3. These shifts, attributed to temperature variations in optical fibers altering the optical delay or positioning changes in the delay line, lead to misalignment of the main pulse in each time trace. As a result, when these traces are averaged, noise is introduced, as depicted in Fig. 3. This issue of signal alignment is quite general [41] and not unique to THz-TDS experiments; similar challenges exist in other fields such as sampling oscilloscopes [40, 41], seismology [42, 43], radar [44], and ultrasound for biology [45]. While several methods have been proposed to address this problem, we initially chose to explore the simplest solutions.

To correct this misalignment smaller than the size of a time sampling step, we implemented the standard formula of the Fourier transform of a translation. A time trace among the recorded time traces is chosen as a reference and all time traces are readjusted one by one to this time trace before averaging. The reference time trace is defined as the time trace closest to the average:

$$T_{ref} = \arg \min_{T_i \in \mathcal{T}} \frac{\overline{T} \cdot T_i}{\|T_i\|_2}. \quad (6)$$

By naming  $\delta_i$  the delay shift on the time trace  $i$ , the corrected signal in frequency domain is:

$$\tilde{E}_{i-corrected}(f) = \exp(j2\pi f \cdot \delta_i) \cdot \tilde{E}_i(f) \quad (7)$$

$$\delta_i = \arg \min_{\delta_i} \|T_{ref} - T_{i-corrected}(\delta_i)\|_2. \quad (8)$$

The goal in Eq.8 is to find for each time trace the drift which minimize the distance between the reference and the corrected time trace.

Fig. 5 shows the mean and the precision on 1 000 time traces after the drift correction. The signal to noise ratio increase from 28dB to 55dB, which means approximately a gain of 27dB.

Furthermore, averaging shifted time traces is equivalent to applying a convolution low-pass filter to the signal as explained in [39]. Assuming a linear drift during the acquisition for an overall drift  $\Delta$ , for an average over a large number of traces, the attenuation factor in *amplitude* caused by the drift can be approximated in the frequency domain as

$$A_{\Delta}(\omega) = \frac{1}{\Delta} \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} e^{j\omega\delta} d\delta = \text{sinc}\left(\frac{\omega\Delta}{2}\right). \quad (9)$$

More generally, the attenuation is the Fourier transform of the delay distribution among all the time traces. Equation 9

corresponds to the special case of an uniform distribution. It is important to note that the attenuation does not depends directly on the number of traces, but only on the delay distribution. Hence, the simple average of the raw time traces is a biased estimator of the signal, and does not approach the real signal for large number of time traces. In fact, the delay distribution is more likely to be broader for long acquisitions, such that the accuracy of the uncorrected average *decreases* with the number of time traces. Fig. 4 shows the ratio between the corrected average and the raw average for 1000 and 50000 time traces. In this example, the delay drift caused a severe attenuation in the raw average, and gets much worse for a larger number of averaged time traces. The attenuation closely match the form predicted in equation 9 up to 4 THz, despite a non linear drift (see supplementary document for details on the actual drift).

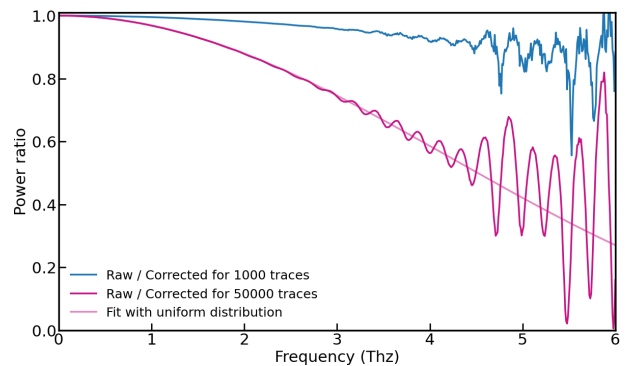


Fig. 4: Power ratio of spectra obtained from a raw average and a corrected average showing the multiplicative bias caused by the drift. Shown for 1000 and 50000 time traces. The drift attenuation for 50000 time traces is fitted with the model from equation 9.

Figure 4 highlights the potentially misleading nature of the error caused by the delay drift. Indeed, despite the signal being affected by more than 20% at 3 THz, the signal does not appear more "noisy", and it is very difficult to distinguish this error from a real signal if only the raw average is taken into consideration. In fact, it could well be an explanation of the dispersion in the published results of THz spectral analysis of water using THz-TDS techniques ([35, 46] and references therein), where the strong absorption makes long accumulations necessary.

#### B. Delay line speed variation

At the view of the major gain obtained by the delay correction at the first order, we decided to go a step further. In fact, the variations in speed of the delay line due to the temperature or any other environmental fluctuation, has already been reported as in [37]. Hence, there can be at the second order, noise due to the phenomena describe above. It leads to a dilatation in the time axis where the measured signal is  $E(t + \alpha t)$  instead of  $E(t)$ . A dilatation in the time domain implies a contraction in the frequency domain, which is a nonlinear transformation. To correct this, we stay in the time domain and use the Taylor formula at first order.



$$E_{i\text{-corrected}}(t) = E_i(t) - (\alpha_i t) E_i'(t) \quad (10)$$

$$\alpha_i = \arg \min_{\alpha_i} \|T_{ref} - T_{i\text{-corrected}}(\alpha_i)\|_2 \quad (11)$$

After the delay line speed correction, the gain is really small and there is almost no difference in the time and frequency domain, when applied to a single peak reference as we are doing here. However, we performed experiments outside of the scope of this paper and realized that this correction becomes important when analyzing time traces containing echoes as the one coming from the recording of a solid state sample (the echoes come from the Fabry Perrot effect).

### C. Laser amplitude fluctuation

The gain after the drift correction is important, however there is still noise remaining. This noise corresponds to other sources such as shot noise and fluctuations in laser power. We suppose this noise is proportional to the signal with a coefficient  $a_i$  for each time trace  $i$ . Hence the corrected signal in time domain is:

$$E_{i\text{-corrected}}(t) = (1 - a_i) E_i(t) \quad (12)$$

$$a_i = \arg \min_{a_i} \|T_{ref} - T_{i\text{-corrected}}(a_i)\|_2 \quad (13)$$

Here the goal is the same as before, find for each time trace the coefficient which minimize the distance between the reference and the corrected time trace.

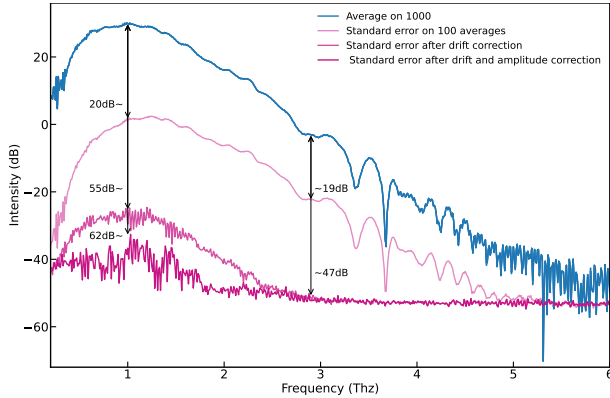


Fig. 5: Mean and standard error in frequency and time domain after drift and amplitude correction

Fig. 5 shows the mean and the precision on 1000 time traces after the drift and the amplitude correction. In this case the signal estimated does not change. However the noise estimated does. Hence, the signal to noise ratio increase from 55 dB to 62 dB, which means approximately 7 dB gain.

### D. Periodic sampling errors

When averaging repeating measures, it appears that there is a distortion in the spectrum, which makes the spectrum of the estimated signal unusable and can significantly influence the extraction of the refractive index of the sample at high frequencies (typically  $> 7.5$  THz). Fig. 6 shows the mean on

50 000 time traces before correction in gray and after drift plus other corrections in blue. After approximately 7.5 THz there is an artifact and between 4.5 THz and 7.5 THz, the artifact is present but hidden by the signal intensity, this area is then not reliable depending of the number of time traces accumulated. As explained in [38], this distortion is called periodic sampling error and comes from periodic deviation of the delay line position. It leads to apparently higher artificial bandwidth due to parasitic mirror copies of the main pulse spectrum around the error's frequency. Indeed, at each time  $t$ , instead of measuring the signal  $E(t)$  the system measures  $E(t + \sigma(t))$ ,  $\sigma(t)$  being a periodic error in the time domain. It is possible to recover an approximation of the normal spectrum from the measured spectrum, by minimizing the electric field in the concerned area. Indeed there is a frequency we call  $f_{ps} \in \{f_0, \dots, f_{\lfloor \frac{p-1}{2} \rfloor}\}$  that can be identify graphically, where the error start. The corrected signal is:

$$E_{i\text{-corrected}}(t) = E_i(t) - E_i'(t) \cdot \sigma_i(t) \quad (14)$$

$$\sigma_i(t) = A_i \cos(\nu_i t + \phi_i) \quad (15)$$

$$(A_i, \nu_i, \phi_i) = \arg \min_{A_i, \nu_i, \phi_i} \sum_{f \in f_{ps}, \dots, f_{\lfloor \frac{p-1}{2} \rfloor}} |\tilde{E}_i(f)| \quad (16)$$

$\sigma(t)$  is found by minimizing the integral of the spectrum in the concerned area.  $f_{ps}$  is approximately 7.5 THz in Fig. 6 showing this noise and its reduction by the proposed method.

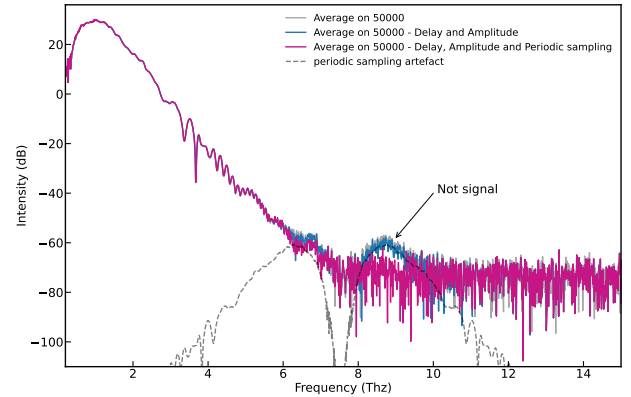


Fig. 6: Spectrum with periodic sampling error beginning at 7.5 THz before and after correction with a Tukey window ( $\alpha = 0.05$ )

In this section, we proposed and implemented in Correct@TDS two important corrections for the signal estimator in THz-TDS. As in sampling oscilloscope for instance [40], the time sampling drift lead to low-pass effect on the amplitude characteristic that can be even more detrimental here because if its effect on the reference would be more important than on the sample this would lead to an artificial gain measurements on the sample. We anticipate that, as in fast electronics and ultrasonic systems[47], further corrections in both the systems and in the data processing would further improve the signal estimator in the future. This will be even more interesting and impactful regarding the new developed THz-TDS ultra-rapid systems. There the mechanical delay line is replaced by

optoelectronic elements as in asynchronous optical sampling systems (ASOPS) [48, 49], or by two coupled laser as dual comb spectroscopy [50]. Despite their speed to acquire a spectrum and their high resolution these systems are facing issues to reach as good signal to noise ratio as the mechanical systems for the same integration time and will surely benefit of specific and more sophisticated noise correction and signal estimators.

#### IV. MATERIAL PARAMETERS EXTRACTION FROM ESTIMATED SIGNALS

Spectroscopy studies how a sample, a material, behaves when crossed by an electromagnetic wave. The goal is to extract parameters describing the physics of the electric charges in the sample. In THz-TDS, this extraction requires the knowledge of the THz pulse before and after the sample, which translate in without and with the sample. The measurements without the sample is called the reference pulse and the one with the sample the sample pulse. Once recorded both pulses are processed to estimate both the signal and the noise.

##### A. Fitting process

Then to go further in the interpretation of the measurements the physicist select a permittivity model set in a transfer function transforming the reference pulse into the sample pulse. Since most of the dielectric function models are depicted in the frequency domain, most of the fits are performed there. However, several groups of which we are part have decided to perform it in the time domain to be as close as possible to the experimental signal. In [27], we implement our approach in a software called called Fit@TDS where the cost function to compare a model with its set of parameters  $P$  to the experiments, is defined as the squared error between the sample pulse and the model fitted curve:

$$C_{model}\{P\} = \left\| E_{sample} - E_{model}\{P\} \right\|_2 \quad (17)$$

$$E_{model}\{P\} = f(P, E_{ref}) \quad (18)$$

The function  $f$  corresponds to the model, while  $C_{model}\{P\}$  is the error associated with the model and its parameters  $P$ .

Most fitting models implicitly assume that the noise is Gaussian and independent and identically distributed (i.i.d.). However, as shown in Fig. 3, the noise present in THz-TDS signals is not a Gaussian white noise, then its impact cannot be ignored. To address this issue, [26] proposed to incorporate the noise correlation matrix, into the cost function in the fitting process to account for measurement uncertainties. This approach involves normalizing the model cost with the matrix and selecting the best fitting model based on the Akaike criterion, which balances goodness of fit and model complexity in terms of number of parameters [29].

The noise correlation matrix called here  $M_{noise}$  contains the information about the noise in the sample signal, the reference signal and the correlations. It enables a proper weighting of the cost function during the fitting process and therefore improves the accuracy of the material parameter extraction. It leads to a new associated cost function to the fitting model:

$$C_{model}\{P\} = \left\| [M_{noise}]^{-1/2} (E_{sample} - E_{model}\{P\}) \right\|_2 \quad (19)$$

$$M_{noise} = \Sigma_{E_{sample}} + \Sigma_{h*E_{ref}} \quad (20)$$

$$C_{AIC}\{P\} = C_{model}\{P\} + 2N_P \quad (21)$$

$C_{AIC}\{P\}$  is the Akaike criterion associated with the model and its parameters  $P$ ,  $N_P$  is the number of parameters of the model. The equation (20) represents the noise correlation matrix as a sum of two covariance matrices. The first term,  $\Sigma_{E_{sample}}$ , is the covariance matrix of the sample measurements. The second term,  $\Sigma_{h*E_{ref}}$ , is the covariance matrix of the reference measurements, convolved with the transfer function  $h$  of the THz-TDS system. The transfer function  $h$  accounts for the distortion of the THz pulse as it propagates through the system. By convolving the reference with the transfer function, the noise correlation matrix of the reference is transferred to the sample one and added to the noise correlation matrix of the sample.

In practice, the true values of  $h$  and the covariances  $\Sigma_{E_{sample}}$  and  $\Sigma_{h*E_{ref}}$  are unknown and have to be estimated from measured data. The estimation of the off-diagonal values of the covariance matrices is important because there are often correlations, as we have seen previously. Moreover, these correlations increase in the presence of a sample due to its own variations, making their estimation even more crucial. Therefore, the estimators should be able to approximate these off-diagonal values accurately. One approach is to estimate them from a set of repeated measurements, but there can be errors caused by the system, leading to a high variance in the signal. This problem highlights the importance of our corrections, which reduce the noise and enable a less biased estimation of the covariance matrices.

##### B. Noise correlation matrix

Covariance matrices are used in all field of signal from sampling oscilloscope [51], acoustics [52] and biomedical application [53]. It serves, for instance, to propagate uncertainties, to shrink huge data to a reasonable size, and to obtain the necessary normalization [51] in a fitting process. In our case, we first estimate  $E_{sample}$  and  $E_{ref}$  with the proposed method : averaging  $n_{acc}$  time traces after correcting the delay line and the laser amplitude errors.

$$E_{ref} \approx E_{ref-n_{acc}} = \frac{1}{n_{acc}} \sum_{i=1}^{n_{acc}} T_{i-(ref)} \quad (22)$$

$$E_{sample} \approx E_{sample-n_{acc}} = \frac{1}{n_{acc}} \sum_{i=1}^{n_{acc}} T_{i-(sample)} \quad (23)$$

Following, the empirical value of the transfer function  $h$  is derived as follow:

$$h = \text{TF}^{-1} \left[ \frac{\tilde{E}_{sample}}{\tilde{E}_{ref}} \right] \approx \text{TF}^{-1} \left[ \frac{\tilde{E}_{sample-n_{acc}}}{\tilde{E}_{ref-n_{acc}}} \right] \quad (24)$$

Also, the computation of  $M_{noise}$  requires to know the covariance matrices associated to the sample and the transformed reference. Since  $E_{sample}$  and  $h*E_{ref}$  are estimated by averaging

$n_{acc}$  corrected time traces, their covariance matrices can be estimated using the following formula:

$$\Sigma_{h^*E_{ref}} \approx \frac{1}{n_{acc}} \Sigma_{h^*E_{ref-n_{acc}}} \quad (25)$$

$$\Sigma_{E_{sample}} \approx \frac{1}{n_{acc}} \Sigma_{E_{sample-n_{acc}}} \quad (26)$$

We divide by  $n_{acc}$  in the formula because we are estimating the covariance of the mean of the corrected time traces, not the covariance of each individual trace. When we average the  $n_{acc}$  corrected time traces, we reduce the variance of the resulting mean by a factor of  $n_{acc}$ . Thus, dividing by  $n_{acc}$  in the covariance computation accounts for this reduction in variance due to averaging.

Finally, the normalization in equation (19) implies the computation of the inverse of the noise correlation matrix. If the two covariance matrices in equation (20) are invertible, that means definite positive then their sum is also invertible. But there is no assumption on this invertibility if the two covariance matrices are singulars.

### C. Covariance and precision matrices estimators

A necessary condition for the inversion of an estimated covariance is that the number of sample meaning the number of repeated measurements used for the estimation, must be strictly greater than the number of variables meaning the number of time sampling points. In practice, the acquisition of sufficient number of time traces could be very long and an unpractical when dealing with numerous experimental conditions to test. For example, a time trace of 400 ps is about 12 500 time sampling points, and the acquisition of 12 501 time traces is about 2 hours. Also, the more time traces are accumulated, the more drifts errors are accumulated from intrinsic instability.

The literature offers various methods to approximate well-conditioned covariance matrices. Our goal is to propose a simple, reproducible process for approximating well-conditioned covariance matrices and to integrate this process into our open-source software solution, Correct@TDS, making it easily accessible to users. To achieve this, we utilize methods implemented in the widely-used Python library, Scikit-learn [54]. We examine three covariance estimation methods: empirical covariance, shrinkage methods, and graphical lasso.

The empirical covariance matrix  $\Sigma_{emp}$ , can be unreliable as an estimator when  $n$  is close to or less than  $p$ , leading to ill-conditioned covariance estimates due to the large magnitude gap between the eigenvalues. To avoid this inversion problem, the shrinkage method has been proposed by [55] (Ledoit-Wolf shrinkage). It is a transformation of the empirical covariance matrix toward a target matrix with regularization techniques:

$$\Sigma_{shrunk} = (1 - \alpha)\Sigma_{emp} + \alpha \frac{\text{Tr}(\Sigma_{emp})}{p} I \quad (27)$$

The target matrix considered here is proportional to the weighted sum of the identity matrix and the empirical matrix. The optimal value of  $\alpha$  is computed by finding a bias-variance

trade-off with no assumption on the data distribution. Another methods called Oracle Approximating Shrinkage method [56] had been proposed as an improvement of Ledoit-Wolf shrinkage where  $\alpha$  is computed under Gaussian distribution assumption on the samples.

There are also methods that rely on solving the covariance selection problem to approximate sparse inverse covariance matrices with lasso regularization [57] [58] [59]. These methods have the advantage to be able to recover the off diagonal structure of the covariance matrix. Since two independent variables will have zero coefficient in the precision matrix, the idea is to estimate a sparse covariance matrix of multivariate Gaussian distributed observations by minimizing the penalized negative log-likelihood:

$$\Sigma_{glasso} = \arg \min_{\Sigma} -\log(\det \Sigma) + \text{Tr}(\Sigma_{emp}\Sigma) + \alpha \|\Sigma\|_1 \quad (28)$$

The coefficient  $\alpha$  is the lasso penalty; its value is found by comparing different values on a grid of  $\alpha$  values. The larger alpha is, the more sparse the matrix is. This method tends to work better when  $n \leq p$ , however the lasso penalization based methods can be inconsistent on some graph structures [60]. In [59], they proposed the graphical lasso, a fast algorithm based on coordinate descent approach to solve this problem.

For the sake of the simplicity in the explanations, we limited the scope of the paper to data of a reference, meaning without samples in the estimation of  $\Sigma_{E_{ref-n_{acc}}}$  and its inverse. It worth noticing that the process is the same for  $\Sigma_{E_{sample-n_{acc}}}$  and  $\Sigma_{h^*E_{ref-n_{acc}}}$  and we expect that the noise induced by the signal will be lower simply because a sample will reduce the magnitude of the signal.

The tests are done on the measured time traces on 100 ps in the third experiment in section 2. Since we don't know the distribution of the residual noise in Fig. 5, we tried different estimation methods for different values of  $n$  with  $n \leq p$  and compared them to targets matrices which are supposed to be the true covariance/precision matrices. Those targets matrices are obtained by computing the empirical covariance and its inverse on  $n = 100\,000$  time traces with  $p = 3000$  and  $p/n = 0.03$ . We assume that we have enough data for a well conditioned matrix close to the true matrices.

### D. Results and discussion

Our goal was to compare the different methods to estimate the covariance and precision matrices. On this basis, we tested Ledoit Wolf shrinkage, Oracle Approximating shrinkage and the Graphical Lasso. This section shows the obtained results.

1) *Covariance matrix*: Figure 7(a) illustrates the normalized distance between the estimated covariance matrix for different number of sample and the empirical target covariance matrix. This distance is defined as the Frobenius norm of the difference between the two matrices, normalized by the Frobenius norm of the target matrix, with an expected value of 0. The gray line indicates the point at which  $n = p + 1$ , indicating the invertibility of the covariance matrices. We can see that the two shrinkage methods produce similar results



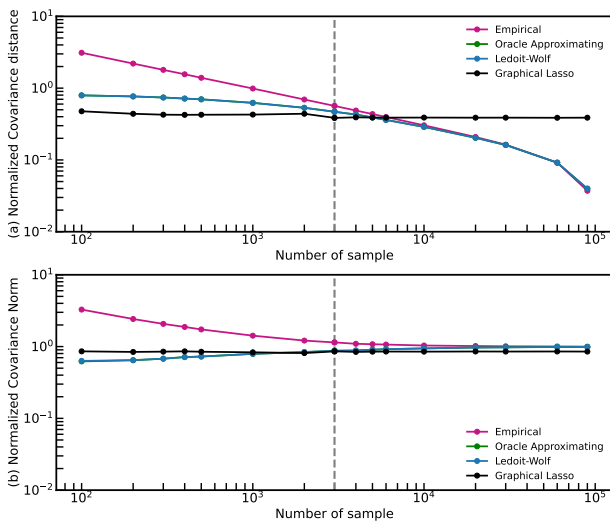


Fig. 7: Comparison between Ledoit Wolf shrinkage, Oracle Approximating shrinkage and the Graphical Lasso methods for the covariance matrix estimation. (a) Normalized distance between estimated and target covariance matrices for different number of sample showing a good estimation with the graphical lasso already with 100 time traces. (b) Normalized norm of the estimated covariance matrices showing that the norm of the estimated matrices are in the range of the targeted ones and that the graphical Lasso performs better

despite their different assumptions about the Gaussian distribution. When  $n \leq p$ , which is the relevant scenario of interest for us, the graphical lasso algorithm performs better even with small number of time traces such as 100.

In addition to the comparison of the distances, we also compare the norm of the estimated matrices to the norm of the target matrices. Figure 7(b) shows the ratio between these two norms, with an expected value of 1. The ratio ranges from approximately 0.6 to 1 depending on the estimation method. These results suggest that the estimated matrices are in close neighborhood of the target matrix and have similar norms, which indicates that the estimation methods produce accurate results.

Moreover, in order to evaluate the accuracy of the estimators in retrieving the covariance matrix structure and correlation values, a plot was generated.

Figure 8(a) displays the empirical covariance matrix before correction on 100 000 time traces. Before correction, the diagonal values are not large compared to the off-diagonal values, indicating a certain degree of correlation between the variables that represent each time point in the signal. The correlations are certainly due to noise or artifacts in the signal and need to be removed or corrected. It should be noted that here, a low frequencies filter (refer section 2) is applied, otherwise the correlations would be stronger. Furthermore, the highest positive and negative covariance values are close to the main pulse, reflecting the position of the main pulse.

In Figure 8(b), the empirical covariance matrix after correction on 100 000 time traces is shown as the target matrix. The correction leads to a reduction of covariance values by

three orders of magnitude, and a change in the diagonal elements, which become larger than the off-diagonal elements. Moreover, the correlation due to the main pulse is still reflected. This suggests that the correlation due to noise or unwanted artifacts in the signal are corrected. Hence we can conclude that the corrections help to obtain an almost accurate representation of the underlying signal.

Figure 9 illustrates the estimated covariance matrix with  $n_{acc} = 1000$  and  $p = 3000$  with Graphical Lasso (a) and Ledoit-Wolf (b). They have the same structure with the target variables, meaning they reflected the main pulse correlation and the large diagonal values. However, the Graphical Lasso algorithm appears to provide a more precise estimation of the main pulse correlation compared to the target covariance matrix, which is consistent with the results shown in Figure 7(a) where the graphical lasso algorithm performed better in terms of distance to the target matrix.

2) *Precision matrix*: The most challenging task is now to determine which estimator is the most precise for retrieving the precision matrix. Obviously we have a curse of dimensionality problem and there is no optimal solution for this. The same comparisons as for the covariance estimator are done for the precision.

Figure 10(a) shows the normalized distances between the estimated precision matrices and the target precision matrix (inverse of the empirical target covariance matrix). The two shrinkage methods also produce here similar results. In the case where  $n \leq p$ , where empirical covariance matrix is not invertible, the distances for all methods are almost in the same order of magnitude. However when we get closer to the invertibility line, the graphical lasso seems to work better.

Regarding the normalized norms of the estimated matrices in Figure 10(b), we can observe that, the norms with the shrinkage methods get closer to the target with more samples while the norms with graphical lasso are almost constant. However all values also range from 0.6 to 1 depending on the estimation method indicating that the estimated matrices are also in close neighborhood of the target matrix.

In Fig. 11, the precision matrix (inverse of the empirical covariance matrix) is shown before and after correction on 100 000 time traces. Despite the number of sample supposed sufficient used for computation, the precision matrix before correction does not reflect the strong correlation on the main pulse as in the covariance, which is contrary to the expectation that two independent variables would have a zero coefficient in the precision matrix, and vice versa. On the other hand, the precision matrix which is our target precision after correction appears to better capture the expected structure. Also there is an increase of the partial correlation values by one order of magnitude after correction.

In Fig. 12, the precision matrices obtained by applying the graphical lasso and Ledoit-Wolf methods on 1000 time traces are presented. Although these matrices appear to be closer to the target matrix (also referring to Fig. 10), the graphical lasso fails to capture the strong correlation on the pulse in the precision matrix. Still, this structure can be explained by the sparsity hypothesis of the precision matrix made in lasso algorithm, can affect the structure and make it different from

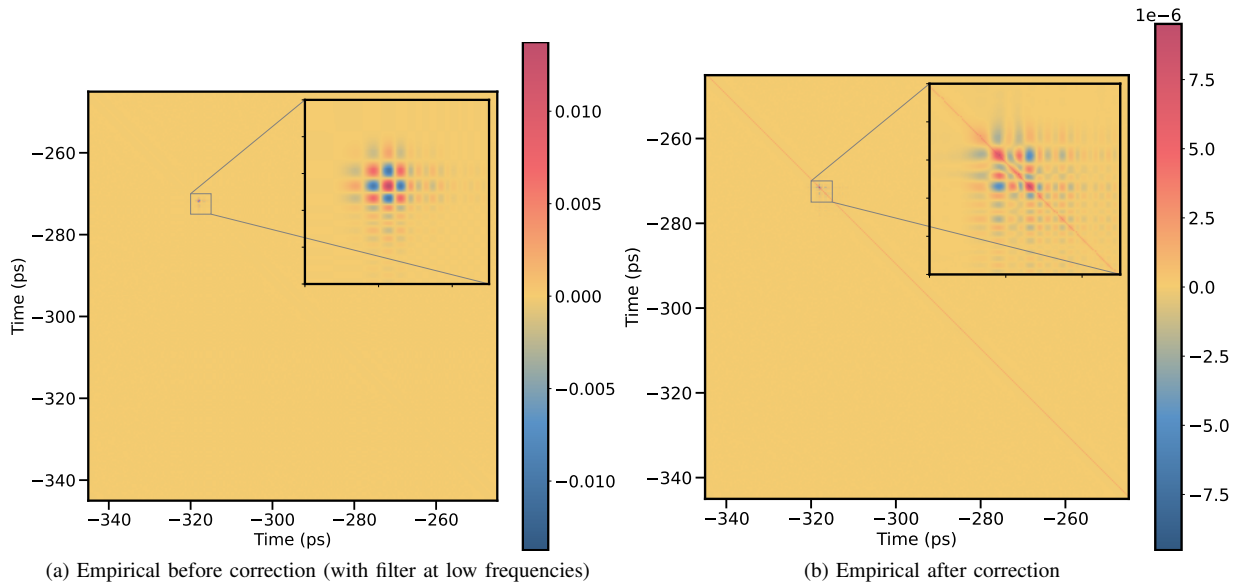


Fig. 8: Empirical covariance before correction and after correction for 100 000 time traces showing the importance of the correction to get a covariance immune to drift

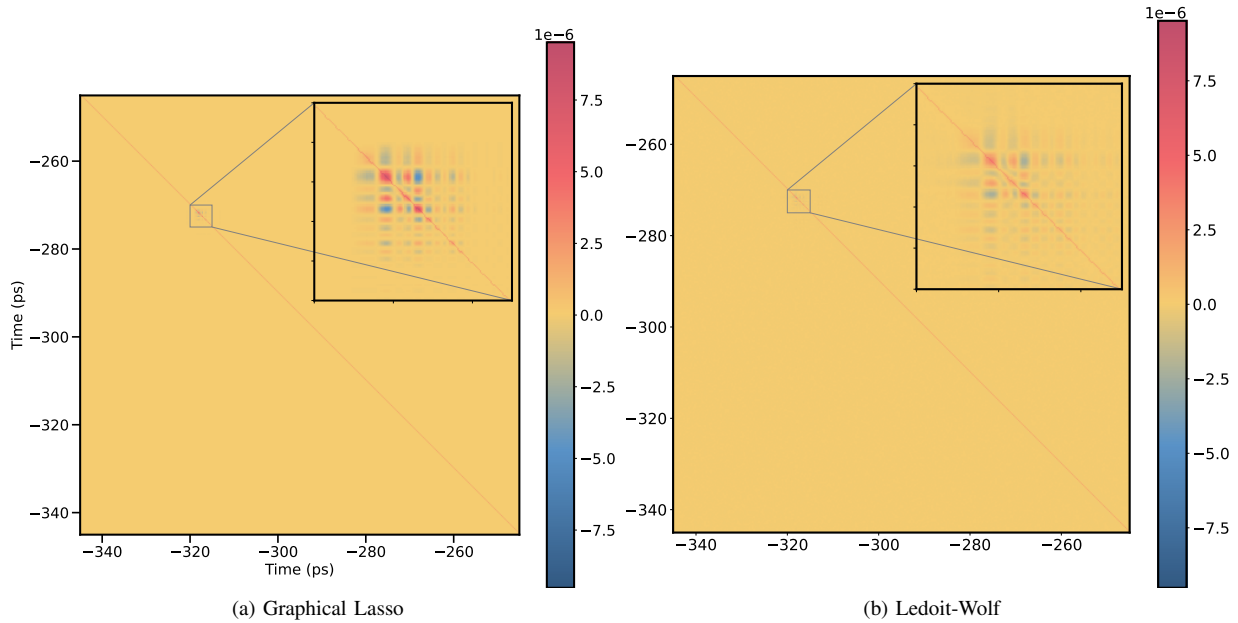


Fig. 9: Estimated Covariance with Graphical Lasso and Ledoit-Wolf algorithms for 1000 time traces

the covariance matrix. As for Ledoit-Wolf, we observe a partial retrieval of the strong correlation, but it remains difficult for the estimators to capture it entirely. It can just be due to the fact that 1 000 time traces is not sufficient for this method.

3) *Best estimator*: Finally, graphical lasso seems to be the best estimator for our problem. The model’s goal is to retrieve the matrix structure. It regularizes the empirical matrix and give sparse precision matrices. While the resulting matrices may not be entirely accurate, they are still suitable for our purpose that is to ensure a proper fit normalization taking into account the noise distribution and therefore giving better evaluation of the extracted physical parameters and not to

get the exact values. Additionally, Graphical lasso has the advantage to give sparse matrix that will be advantageous in terms of computational costs during the fitting process.

Noise correlation matrix estimation is again a general problem in data processing where specific solution can be found for specific problems. Here, we took the first steps in their evaluation in the case of THz-TDS with the goal of ensuring a proper weighting during the fit process giving the parameters interesting for the users of the setup. It leads to the first evaluation of these matrix for this setup to our knowledge with a reasonable accuracy using only python public libraries. We

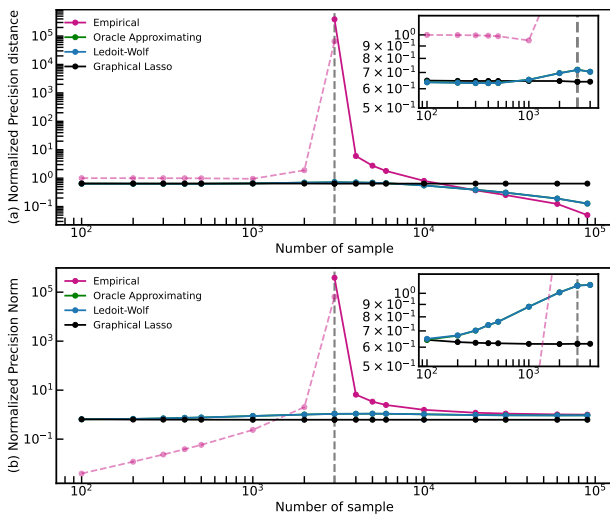


Fig. 10: Comparison between Ledoit Wolf shrinkage, Oracle Approximating shrinkage and the Graphical Lasso methods for the precision matrix estimation. (a) Normalized distance between estimated and target precision matrices for different number of sample showing almost the same results for the three methods when  $n \leq p$ . (b) Normalized norm of the estimated precision matrices showing that the norm of the estimated matrices are in the range of the targeted ones.

have no doubt that more sophisticated methods based on the recent works on the subject [61–63] will give quicker and more accurate evaluation of the matrix. Similarly to what we wrote previously, new systems based on optoelectronics or dual laser are able to record thousands of time traces per second and will benefit a lot from proper sparse evaluation of the noise correlation matrix to avoid to record huge volume of correlated data.

## V. CONCLUSION

In conclusion, we have introduced a nonlinear signal estimator and several techniques for retrieving the noise correlation matrix for THz-TDS experiments. These methods have been implemented in our open-source software, Correct@TDS, readily available to the community. Our signal correction significantly improves the signal-to-noise ratio (over 30 dB enhancement at 1 THz for 1000 accumulated traces). This advancement transforms THz-TDS setups worldwide from capable of measuring small or very attenuated signals to excelling at measuring tiny variations of signal, such as in very thin samples or samples with subtle physical or chemical differences. Additionally, our work brought THz-TDS material parameter extraction to more reliable standards by incorporating the proper noise term as a weighting factor.

Our work opens several exciting perspectives in several areas. This work motivates experimentalists to search for hidden noise sources to further reduce noise even more and approach true Gaussian white noise. We are already beginning the integration of Correct@TDS with our existing software, FIT@TDS [27], to extract error bars for fitted parameters, even with our super-resolution technique [64].

Moreover, the fields of THz-TDS signal processing [65, 66] and parameter retrieval [14, 67, 68] have a rich history of over two decades, primarily led by physicists with a focus on other areas of expertise. We anticipate our work to foster stronger collaborations between the THz and the signal processing communities. The latter could leverage concepts and techniques from speech recognition, seismology, radar, or biomedical echography to build faster, more accurate, and improved data processing methods. The widespread adoption of THz analysis in analytical fields like pharmaceuticals and nondestructive testing relies on continued improvements in measurement reliability and reproducibility, areas where standardized data processing can make significant contributions.

## ACKNOWLEDGMENTS

The authors thanks Claire Mantel for the fruitful discussions, the Région "Hauts-de-France" for the funding of the Diagnostera startair project and the DEUS MAROONER stimule project.

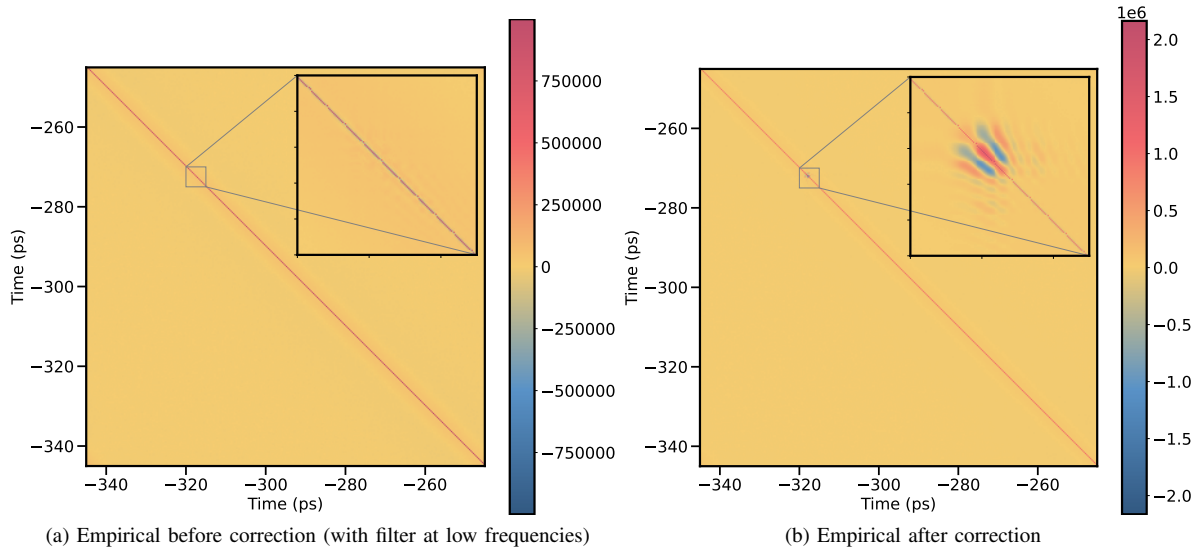


Fig. 11: Empirical Precision before correction and after correction for 100 000 time traces

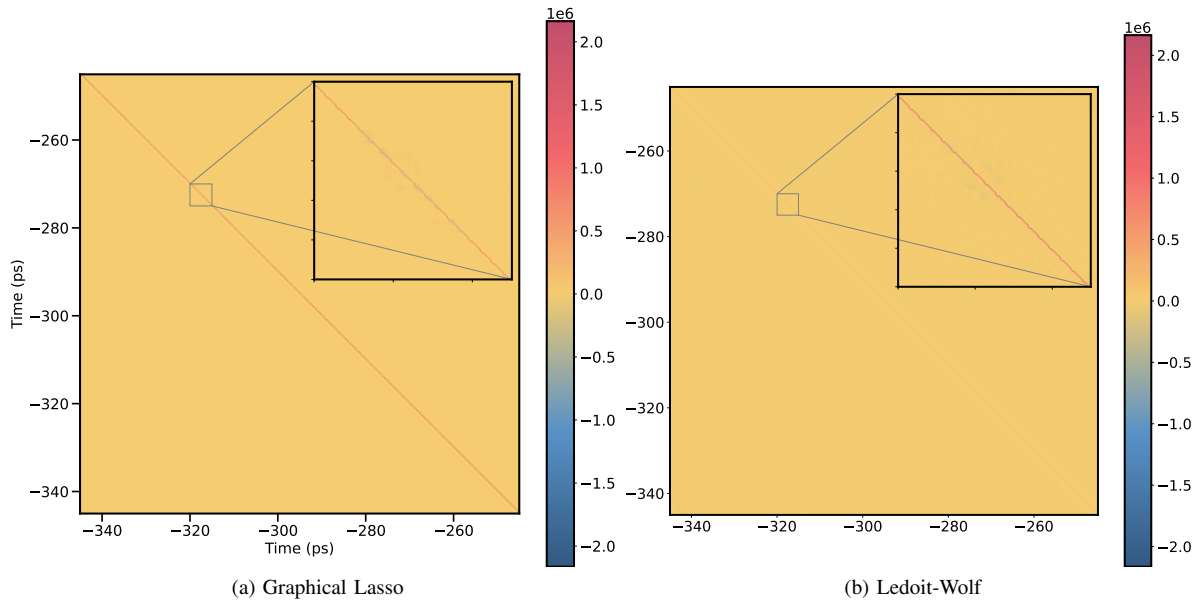


Fig. 12: Estimated Precision with Graphical Lasso and Ledoit-Wolf algorithms for 1000 time traces

REFERENCES

[1] S. Houver, S. E. Barois, P. Roy, and R. Peretti, “La spectroscopie térahertz: électrons et vibrations,” *Photoniques*, no. 121, pp. 36–41, 2023.

[2] R. M. Smith and M. A. Arnold, “Selectivity of terahertz gas-phase spectroscopy,” *Analytical Chemistry*, vol. 87, no. 21, pp. 10 679–10 683, 2015, pMID: 26436438.

[3] E. M. Adams, O. Lampret, B. König, T. Happe, and M. Havenith, “Solvent dynamics play a decisive role in the complex formation of biologically relevant redox proteins,” *Physical Chemistry Chemical Physics*, vol. 22, no. 14, pp. 7451–7459, 2020.

[4] J.-H. Cheng, R. C. Jones, O. Sushko, Y. Tashiro, and R. Donnan, “Quasi-optical sub-thz circular dichroism spectroscopy of solvated myoglobin,” *IEEE Transactions on Terahertz Science and Technology*, vol. 10, no. 4, pp. 348–357, 2020.

[5] M. Failla, J. Keller, G. Scalari, C. Maissen, J. Faist, C. Reichl, W. Wegscheider, O. Newell, D. Leadley, M. Myronov *et al.*, “Terahertz quantum hall effect for spin-split heavy-hole gases in strained ge quantum wells,” *New Journal of Physics*, vol. 18, no. 11, p. 113036, 2016.

[6] S. Biasco, F. Burri, S. Houver, E. Abreu, M. Savoini, and S. L. Johnson, “Impact ionization in low-band-gap semiconductors driven by ultrafast terahertz excitation: Beyond the ballistic regime,” *Physical Review B*, vol. 106, no. 23, p. 235201, 2022.

[7] S. Rajabali, S. Markmann, E. Jöchl, M. Beck, C. A. Lehner, W. Wegscheider, J. Faist, and G. Scalari, “An ultrastrongly coupled single terahertz meta-atom,” *Nature communications*, vol. 13, no. 1, p. 2528, 2022.

[8] T. Dang, J. Hawecker, E. Rongione, G. Baez Flores, D. To, J. Rojas-Sanchez, H. Nong, J. Mangeney, J. Tignon, F. Godel *et al.*, “Ultrafast spin-currents and charge conversion at 3 d-5 d interfaces probed by time-domain terahertz spectroscopy,” *Applied Physics Reviews*, vol. 7, no. 4, p. 041409, 2020.

[9] T. S. Seifert, N. M. Tran, O. Gueckstock, S. M. Rouzegar, L. Nadvornik, S. Jaiswal, G. Jakob, V. V. Temnov, M. Muenzenberg, M. Wolf *et al.*, “Terahertz spectroscopy for all-optical spintronic characterization of the spin-hall-effect metals pt, w and cu80ir20,” *Journal of Physics D:*

- Applied Physics*, vol. 51, no. 36, p. 364003, 2018.
- [10] F. Valmorra, G. Scalari, C. Maissen, W. Fu, C. Schönerberger, J. W. Choi, H. G. Park, M. Beck, and J. Faist, “Low-bias active control of terahertz waves by coupling large-area cvd graphene to a terahertz metamaterial,” *Nano letters*, vol. 13, no. 7, pp. 3193–3198, 2013.
- [11] J.-C. Deinert, D. Alcaraz Iranzo, R. Pérez, X. Jia, H. A. Hafez, I. Ilyakov, N. Awari, M. Chen, M. Bawatna, A. N. Ponomaryov *et al.*, “Grating-graphene metamaterial as a platform for terahertz nonlinear photonics,” *ACS nano*, vol. 15, no. 1, pp. 1145–1154, 2020.
- [12] S. Mityukovskiy, D. E. Vanpoucke, Y. Bai, T. Hannotte, M. Lavancier, D. Hourlier, G. Roos, and R. Peretti, “On the influence of water on the vibrational spectral features of molecular crystals,” *Physical Chemistry Chemical Physics*, vol. 24, no. 10, pp. 6107–6125, 2022.
- [13] P. Bawuah and J. A. Zeidler, “Advances in terahertz time-domain spectroscopy of pharmaceutical solids: A review,” *TrAC Trends in Analytical Chemistry*, vol. 139, p. 116272, 2021.
- [14] L. Duvillaret, F. Garet, and J.-L. Coutaz, “A reliable method for extraction of material parameters in terahertz time-domain spectroscopy,” *IEEE Journal of selected topics in quantum electronics*, vol. 2, no. 3, pp. 739–746, 1996.
- [15] —, “Highly precise determination of optical constants and sample thickness in terahertz time-domain spectroscopy,” *Applied optics*, vol. 38, no. 2, pp. 409–415, 1999.
- [16] Z. Jiang, M. Li, and X.-C. Zhang, “Dielectric constant measurement of thin films by differential time-domain spectroscopy,” *Applied Physics Letters*, vol. 76, no. 22, pp. 3221–3223, 2000.
- [17] T. D. Dorney, R. G. Baraniuk, and D. M. Mittleman, “Material parameter estimation with terahertz time-domain spectroscopy,” *JOSA A*, vol. 18, no. 7, pp. 1562–1571, 2001.
- [18] U. Tayvah, J. A. Spies, J. Neu, and C. A. Schmuttenmaer, “Nelly: A user-friendly and open-source implementation of tree-based complex refractive index analysis for terahertz spectroscopy,” *Analytical Chemistry*, vol. 93, no. 32, pp. 11 243–11 250, 2021.
- [19] J. Liao, B. Wang, Z. Wang, and L. Zhu, “Amino-acid classification based on terahertz absorption spectroscopy with gaussian process and maximum likelihood,” *Sensors and Actuators B: Chemical*, p. 133806, 2023.
- [20] W. Wang, N. Zhu, Z. Wang, C. Zhao, Z. Song, X. Chen, and X. Chao, “Efficient terahertz absorption gas sensor with gaussian process regression in time- and frequency-domain,” *Sensors and Actuators B: Chemical*, vol. 369, p. 132349, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925400522009911>
- [21] L. Mohtashemi, P. Westlund, D. G. Sahota, G. B. Lea, I. Bushfield, P. Mousavi, and J. S. Dodge, “Maximum-likelihood parameter estimation in terahertz time-domain spectroscopy,” *Opt. Express*, vol. 29, no. 4, pp. 4912–4926, Feb 2021. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-29-4-4912>
- [22] J. Lee, C. K. Leung, M. Ma, J. Ward-Berry, S. Santitewagun, and J. A. Zeidler, “The dotthz project: A standard data format for terahertz time-domain data,” *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 44, no. 11, pp. 795–813, 2023.
- [23] E. Denakpo, H. Théo, N. Osseiran, and R. Peretti, “Software: Correct@tds,” 07 2024. [Online]. Available: <https://github.com/THzbiophotonics/Correct-TDS>
- [24] E. Héroult, F. Garet, and J.-L. Coutaz, “On the possibility of identifying substances by remote active thz spectroscopy,” *IEEE Transactions on Terahertz Science and Technology*, vol. 6, no. 1, pp. 12–19, 2015.
- [25] M. Bernier, F. Garet, J. L. Coutaz, H. Minamide, and A. Sato, “Accurate characterization of resonant samples in the terahertz regime through a technique combining time-domain spectroscopy and kramers-kronig analysis,” *IEEE Transactions on Terahertz Science and Technology*, vol. 6, no. 3, pp. 442–450, 2016.
- [26] L. Mohtashemi, P. Westlund, D. G. Sahota, G. B. Lea, I. Bushfield, P. Mousavi, and J. S. Dodge, “Maximum-likelihood parameter estimation in terahertz time-domain spectroscopy,” *Optics Express*, vol. 29, no. 4, pp. 4912–4926, 2021.
- [27] R. Peretti, S. Mityukovskiy, K. Froberger, M. A. Mebarki, S. Eliet, M. Vanwollegem, and J.-F. Lampin, “Thz-tds time-trace analysis for the extraction of material and metamaterial parameters,” *IEEE transactions on Terahertz Science and Technology*, vol. 9, no. 2, pp. 136–149, 2018.
- [28] M. Lavancier, N. V. Yassine, J. Vlieghe, J.-F. Lampin, and R. Peretti, “A criterion to compare permittivity models in the terahertz range,” in *2021 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz)*. IEEE, 2021, pp. 1–2.
- [29] M. Lavancier, N. Vindas-Yassine, J. Vlieghe, T. Hannotte, J.-F. Lampin, F. Orioux, and R. Peretti, “A new metric for the comparison of permittivity models in terahertz time-domain spectroscopy,” *IEEE Transactions on Terahertz Science and Technology*, pp. 1–10, 2024.
- [30] A. G. Markelz and D. M. Mittleman, “Perspective on terahertz applications in bioscience and biotechnology,” *ACS Photonics*, vol. 9, no. 4, pp. 1117–1126, 2022. [Online]. Available: <https://doi.org/10.1021/acsp Photonics.2c00228>
- [31] L. Duvillaret, F. Garet, and J.-L. Coutaz, “Influence of noise on the characterization of materials by terahertz time-domain spectroscopy,” *JOSA B*, vol. 17, no. 3, pp. 452–461, 2000.
- [32] W. Withayachumankul, B. M. Fischer, H. Lin, and D. Abbott, “Uncertainty in terahertz time-domain spectroscopy measurement,” *JOSA B*, vol. 25, no. 6, pp. 1059–1072, Jun 2008. [Online]. Available: <https://opg.optica.org/josab/abstract.cfm?URI=josab-25-6-1059>
- [33] M. Naftaly, “Metrology issues and solutions in thz time-domain spectroscopy: Noise, errors, calibration,” *IEEE Sensors Journal*, vol. 13, no. 1, pp. 8–17, 2012.
- [34] C. Muller, M. Lavancier, J. Bichon, T. Hannotte, J.-F. Lampin, S. Eliet, and R. Peretti, “Beyond the” dynamic range” approach in noise evaluation for terahertz time domain spectrometers,” in *2021 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz)*. IEEE, 2021, pp. 1–2.
- [35] M. Lavancier, “Heuristic approach to take up the challenge of terahertz time-domain spectroscopy for biology,” Ph.D. dissertation, Université de Lille, 2021.
- [36] E. Denakpo, T. Hannotte, M. Lavancier, S. E. Barois, and R. Peretti, “Noise analysis, noise reduction and covariance estimation for time domain spectroscopy,” in *2022 47th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz)*. IEEE, 2022, pp. 1–2.
- [37] D. Jahn, S. Lippert, M. Bisi, L. Oberto, J. Balzer, and M. Koch, “On the influence of delay line uncertainty in thz time-domain spectroscopy,” *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 37, no. 6, pp. 605–613, 2016.
- [38] A. Rehn, D. Jahn, J. C. Balzer, and M. Koch, “Periodic sampling errors in terahertz time-domain measurements,” *Optics Express*, vol. 25, no. 6, pp. 6712–6724, 2017.
- [39] T. M. Souders, D. R. Flach, C. Hagwood, and G. Yang, “The effects of timing jitter in sampling systems,” in *6th IEEE Conference Record, Instrumentation and Measurement Technology Conference*, vol. 39, no. 1, IEEE. Institute of Electrical and Electronics Engineers (IEEE), 1989, pp. 199–203.
- [40] F. Verbeyst, Y. Rolain, J. Schoukens, and R. Pintelon, “System identification approach applied to jitter estimation,” in *2006 IEEE Instrumentation and Measurement Technology Conference Proceedings*, IEEE. IEEE, 2006, pp. 1752–1757.
- [41] K. Coakley and P. Hale, “Alignment of noisy signals,” *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 1, pp. 141–149, 2001.
- [42] D. B. Harris, “A waveform correlation method for identifying quarry explosions,” *Bulletin of the Seismological Society of America*, vol. 81, no. 6, pp. 2395–2418, 1991.
- [43] F. Romano, A. Piatanesi, S. Lorito, C. Tolomei, S. Atzori, and S. Murphy, “Optimal time alignment of tide-gauge tsunami waveforms in nonlinear inversions: Application to the 2015 illapel (chile) earthquake,” *Geophysical Research Letters*, vol. 43, no. 21, pp. 11–226, 2016.
- [44] M. Wang, X. Li, L. Gao, Z. Sun, G. Cui, and T. S. Yeo, “Signal accumulation method for high-speed maneuvering target detection using airborne coherent mimo radar,” *IEEE Transactions on Signal Processing*, vol. 71, pp. 2336–2351, 2023.
- [45] R. Al Mukaddim, N. H. Meshram, and T. Varghese, “Locally optimized correlation-guided bayesian adaptive regularization for ultrasound strain imaging,” *Physics in Medicine & Biology*, vol. 65, no. 6, p. 065008, 2020.
- [46] O. Smolyanskaya, N. Chernomyrdin, A. Konovko, K. Zaytsev, I. Ozheredov, O. Cherkasova, M. Nazarov, J.-P. Guillet, S. Kozlov, Y. V. Kistenev, J.-L. Coutaz, P. Mounaix, V. Vaks, J.-H. Son, H. Cheon, V. Wallace, Y. Feldman, I. Popov, A. Yaroslavsky, A. Shkurinov, and V. Tuchin, “Terahertz biophotonics as a tool for studies of dielectric and spectral properties of biological tissues and liquids,” *Progress in Quantum Electronics*, vol. 62, pp. 1–77, 2018.
- [47] X. Gao, Y. Shi, Q. Zhu, Z. Li, H. Sun, Z. Yao, and W. Zhang, “Domain adaptation in intelligent ultrasonic logging tool: From microseismic to pulse-echo,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.
- [48] O. Kliebisch, D. C. Heinecke, and T. Dekorsy, “Ultrafast time-domain spectroscopy system using 10 ghz asynchronous optical sampling with 100 khz scan rate,” *Opt. Express*, vol. 24, no. 26, pp. 29930–29940, 2016. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-24-26-29930>

- [49] M. Oeri, O. Peters, M. Wolferstetter, and R. Holzwarth, "Compact, high-speed sampling engine for pulsed femtosecond lasers," in *Terahertz Photonics*, M. Jarrahi, D. Turchinovich, and S. Preu, Eds., vol. 11348, International Society for Optics and Photonics. SPIE, 2020, p. 1134808.
- [50] H. Fu, X. Jiang, J. Wu, L. Qiu, Y. Yuan, X. Guo, and Y. Zhu, "Terahertz dual-comb spectroscopy: A comparison between time-and frequency-domain operation modes," *Infrared Physics & Technology*, vol. 115, p. 103699, 2021.
- [51] D. Humphreys, P. Harris, and J. Miall, "Instrument related structure in covariance matrices used for uncertainty propagation," in *2012 42nd European Microwave Conference*, IEEE. IEEE, 2012, pp. 1304–1307.
- [52] Y. Yu, Q. Xiong, Z.-S. Ye, X. Liu, Q. Li, and K. Wang, "A review on acoustic reconstruction of temperature profiles: From time measurement to reconstruction algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–24, 2022.
- [53] R. Ali, A. V. Telichko, H. Wang, U. K. Sukumar, J. G. Vilches-Moure, R. Paulmurugan, and J. J. Dahl, "Local sound speed estimation for pulse-echo ultrasound in layered media," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 69, no. 2, pp. 500–511, 2021.
- [54] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [55] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of multivariate analysis*, vol. 88, no. 2, pp. 365–411, 2004.
- [56] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, "Shrinkage algorithms for mmse covariance estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5016–5029, 2010.
- [57] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the lasso," *The Annals of Statistics*, 2006.
- [58] M. Yuan and Y. Lin, "Model selection and estimation in the gaussian graphical model," *Biometrika*, vol. 94, no. 1, pp. 19–35, 2007.
- [59] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.
- [60] O. Heinävaara, J. Leppä-Aho, J. Corander, and A. Honkela, "On the inconsistency of l1-penalised sparse precision matrix estimation," *BMC bioinformatics*, vol. 17, no. 16, pp. 99–107, 2016.
- [61] E. Ollila, D. P. Palomar, and F. Pascal, "Shrinking the eigenvalues of m-estimators of covariance matrix," *IEEE Transactions on Signal Processing*, vol. 69, pp. 256–269, 2020.
- [62] Q. Wei and Z. Zhao, "Large covariance matrix estimation with oracle statistical rate via majorization-minimization," *IEEE Transactions on Signal Processing*, vol. 71, pp. 3328–3342, 2023.
- [63] E. Ollila and A. Breloy, "Regularized tapered sample covariance matrix," *IEEE Transactions on Signal Processing*, vol. 70, pp. 2306–2320, 2022.
- [64] S. Eliet, A. Cuisset, F. Hindle, J.-F. Lampin, and R. Peretti, "Broadband super-resolution terahertz time-domain spectroscopy applied to gas analysis," *IEEE Transactions on Terahertz Science and Technology*, vol. 12, no. 1, pp. 75–80, 2022.
- [65] W. Withayachumnankul and M. Naftaly, "Fundamentals of measurement in terahertz time-domain spectroscopy," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 35, pp. 610–637, 2014.
- [66] D. M. Mittleman, R. H. Jacobsen, R. Neelamani, R. G. Baraniuk, and M. C. Nuss, "Gas sensing using terahertz time-domain spectroscopy," *Applied Physics B*, vol. 67, no. 3, pp. 379–390, 1998.
- [67] T. D. Dorney, R. G. Baraniuk, and D. M. Mittleman, "Material parameter estimation with terahertz time-domain spectroscopy," *JOSA A*, vol. 18, no. 7, pp. 1562–1571, 2001.
- [68] I. Pupeza, R. Wilk, and M. Koch, "Highly accurate optical material parameter determination with thz time-domain spectroscopy," *Optics express*, vol. 15, no. 7, pp. 4335–4350, 2007.